structure is the excessive volume of data stored. Each grid cell will contain more than 1 or 0 values from more than one coverage and a large number of coverages are needed to store different types of information.

MAP model:
This type of data structure integrates the two structure discussed previously. In this raster structure, each thematic coverage is recorded and accessed separately by map name or title. This is accomplished by recording each variable, or mapping unit, of the coverage's theme as a separate number code or label, which can be accessed individually when the coverage is retrieved. The label corresponds to a portion of the legend has its own symbol assigned to it. This structure facilitates the performance of operations on individual grid cells and groups of similar cells, and the resulting changes in value require rewriting only a single number per mapping unit, simplifying the computations. The MAP data structure allows the manipulation of information in a many-to-one relationship of the attribute values and the sets of grids. The MAP is used in GIS mostly.

3. With a neat sketch describe raster and vector data representation.

Answer:
Raster
A raster data type is, in essence, any type of digital image represented by reducible and enlargeable grids. Anyone who is familiar with digital photography will recognize the Raster graphics pixel as the smallest individual grid unit building block of an image, usually not readily identified as an artifact shape until an image is produced on a very large scale. A combination of the pixels making up an image color formation scheme will compose details of an image, as is distinct from the commonly used points, lines, and polygon area location symbols of scalable vector graphics as the basis of the vector model of area attribute rendering. While a digital image is concerned with its output blending together its grid based details as an identifiable representation of reality, in a photograph or art image transferred into a computer, the raster data type will reflect a digitized abstraction of reality dealt with by grid populating tones or objects, quantities, cojoined or open boundaries, and map relief schemas. Aerial photos are one commonly used form of raster data, with one primary purpose in mind: to display a detailed image on a map area, or for the purposes of rendering its identifiable objects by digitization. Additional raster data sets used by a GIS will contain information regarding elevation, a digital elevation model, or reflectance of a particular wavelength of light, Landsat, or other electromagnetic spectrum indicators.

Digital elevation model, map (image), and vector data
Raster data type consists of rows and columns of cells, with each cell storing a single value. Raster data can be images (raster images) with each pixel (or cell) containing a color value. Additional values recorded for each cell may be a discrete value, such as land use, a continuous value, such as temperature, or a null value if no data is available. While a raster cell stores a single value, it can be extended by using raster bands to represent RGB (red, green, blue) colors, colormaps (a mapping between a

thematic code and RGB value), or an extended attribute table with one row for each unique cell value. The resolution of the raster data set is its cell width in ground units.

Raster data is stored in various formats; from a standard file-based structure of TIF, JPEG, etc. to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes. Database storage, when properly indexed, typically allows for quicker retrieval of the raster data but can require storage of millions of significantly sized records.

Vector

In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. Different geographical features are expressed by different types of geometry:

Points

A simple vector map, using each of the vector elements: points for wells, lines for rivers, and a polygon for the lake.

Zero-dimensional points are used for geographical features that can best be expressed by a single point reference — in other words, by simple location. Examples include wells, peaks, features of interest, and trailheads. Points convey the least amount of information of these file types. Points can also be used to represent areas when displayed at a small scale. For example, cities on a map of the world might be represented by points rather than polygons. No measurements are possible with point features.

Lines or polylines

One-dimensional lines or polylines are used for linear features such as rivers, roads, railroads, trails, and topographic lines. Again, as with point features, linear features displayed at a small scale will be represented as linear features rather than as a polygon. Line features can measure distance.

Polygons

Two-dimensional polygons are used for geographical features that cover a particular area of the earth's surface. Such features may include lakes, park boundaries, buildings, city boundaries, or land uses. Polygons convey the most amount of information of the file types. Polygon features can measure perimeter and area.

Each of these geometries is linked to a row in a database that describes their attributes. For example, a database that describes lakes may contain a lake's depth, water quality, pollution level. This information can be used to make a map to describe a particular attribute of the dataset. For example, lakes could be coloured depending on level of pollution. Different geometries can also be compared. For example, the GIS could be used to identify all wells (point geometry) that are within one kilometer of a lake (polygon geometry) that has a high level of pollution.

Vector features can be made to respect spatial integrity through the application of topology rules such as 'polygons must not overlap'. Vector data can also be used to represent continuously varying phenomena. Contour lines and triangulated irregular networks (TIN) are used to represent elevation or other continuously changing values. TINs record values at point locations, which are connected by lines to form an irregular mesh of triangles. The face of the triangles represent the terrain surface.

4. Discuss various types Vector GIS models

Answer: Vector data structures allow the representation of geographic space in an intuitive way reminiscent of the familiar analog map. The geographic space can be represented by the spatial location of items or attributes which are stored in another file for later access. Like raster spatial data model, there are many potential vector data models that can be used to store the geometric representation of entities in the computer.

A point is the simplest spatial entity that can be represented in the vector world with topology. A point requires to be topologically correct with represent to a geographical reference system which locates it with respect to other special entities. To have topology a line entity must consist of an order set of number points with a defined start and end points. Knowledge of the start and end points gives a line direction. For the creation of topologically correct area entities, the data about the points and lines used in its construction, and knowledge of how these are connected to define the boundary, are required.

There are several ways in which vector data structures can be put together into a vector data model by which the relationship between variables in a single coverage or among variable in different coverages can be defined. The two basic types of vector data models are

- Spaghetti Model
- Topological model.

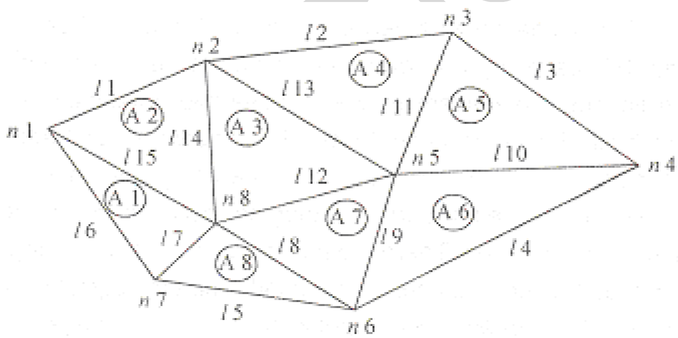
Spaghetti Model:
The simplest vector data structure that can be used to reproduce a geographical image in the computer is a file containing (x,y) coordinate pair that represent the location of individual point features. Fig is essentially a one –for-one translation of the graphical image or a map which is also termed as the conceptual model. Let us consider a conceptual model in which an analog map covering each graphic object is shown in fig. Each graphic object can be represented with a piece of spaghetti. Each piece of spaghetti acts as a single entity. The shortest spaghetti can be represented as a point, collection of a number of point spaghettis for a line entity and collections of line segments that come together at the beginning and ending of surrounding areas from an area entity. Each entity is a single, location recorded in the computer coded as variable length strings of (x,y) coordinate pairs. Let us assume that two polygons lie adjacent to each other in a thematic coverage. These two adjacent polygons must have separate pieces of spaghetti. Each side of polygon is uniquely defined by its own set of lines and

coordinate pairs. In this model of representing vector data, all the spaghettis are recorded separately for polygons. But in the computer they should have the same coordinates.

Topological Models:
In order to use the data manipulation and analysis subsystem more efficiently and obtained the desired result, to allow advanced analytical techniques on GIS data and its systematic study in any project area, much explicit spatial information is to be created. The topological data model incorporates solutions to some of the frequently used operations in advanced GIS analytical techniques. This done by explicitly recording adjacency informing into the basic logical entity in topological data structures, beginning and ending when it contacts or intersects another line, or when there is a change in the direction of the line. Each line then has two sets of numbers: a pair of coordinates and an associated node number. The node is the intersection of two or more lines, and its number is to any line to which it is connected. In addition, each line segment, called a link , has its own identification number that is used as a pointer to indicate the set of nodes that represent its beginning and ending polygon. These links also have identification codes that relate polygon number to see which two polygons are adjacent to each other along its length. In fact the left and right polygons arte also stored explicitly so that even this tedious step is eliminated. This design features allow the computer to know the actual relationship among all its graphical parts to identify the special relationships contained in an analog map document.

There are a number of topological vector data models. Out of the available models there is very common use. These are models are:

- GBF/DIME model created by US department of commerce 1969.
- TIGER model 1986.
- POLYVERT 1984.



Topological vector Data Model.

5. a. Describe the storage of vector data and attribute data.
b. Write short notes on Raster data (10+6)

Answer:

(a): Vector data storage: Vector methods may impose subjective and inexact structure on the landscape, but are more suited to situations where there is a need for precise co-ordinates storage. Important topological information may also be encoded which is very hard to record using raster data structures. The recording of socioeconomic phenomena has generally employed vector techniques, due to the precise nature of the boundaries used for (e.g.) census area definition. A difficulty arises because the precise encoding relates only to the boundaries themselves, and not to the phenomena on which they have been imposed.

In vector representations, an explicit definition is made between the locations of the special entities and the non-spatial attributes of these entities. As mentioned above, these two types of characteristics are frequently held in separate database structures, although some recent work has sought to remove this conventional distinction by database integration. One of the major turnkey GIS packages, ARC/INFO, actually comprises ARC, a special database and manipulation package, and INFO, a commercial DBMS. The independence of the two sub systems is further illustrated by the ability to make alternative software combinations such as ARC/ORACLE. Other turnkey systems, such as Genamap, were designed to provide special data manipulation power which may be built on to an existing relational DBMS (RDBMS), using Structured Query Language (SQL) interfaces (Ingram and Phillips 1987)

(b): Raster data: Raster data is a grid of cells covering an area of interest. Each pixel, the smallest unit of information in the grid, displays a unique attribute.

An example of raster data is a scanned image or photograph. A line drawn in a raster format must be defined by a group of pixels along the length of the line. As a result the size of a raster file is larger than that required by a vector file.

Raster Representation Of Data:

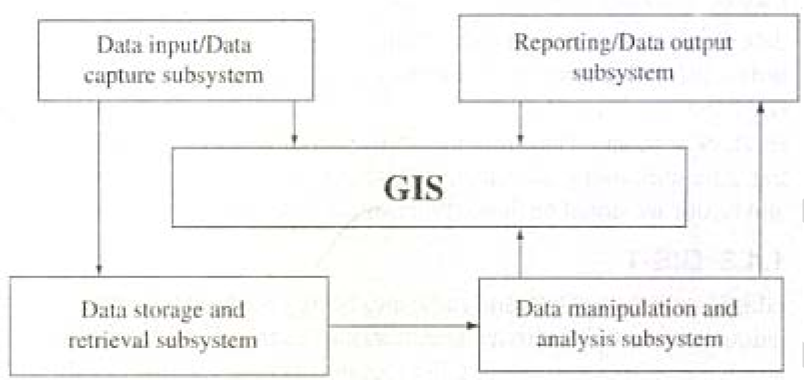
Raster is a method for the storage, processing and display of spatial data. Each area is divided into rows and columns, which form a regular grid structure. Each cell must be rectangular in shape, but not necessarily square. Each cell within this matrix contains location co-ordinates as well as an attribute value. The spatial location of each cell is implicitly contained within the ordering of the matrix, unlike a vector structure which stores topology explicitly. Areas containing the same attribute value are recognised as such, however, raster structures cannot identify the boundaries of such areas as polygons.

Raster data is an abstraction of the real world where spatial data is expressed as a matrix of cells or pixels with spatial position implicit in the ordering of the pixels. With the raster data model, spatial data is not continuous but divided into discrete units. This makes raster data particularly suitable for certain types of spatial operation, for example overlays or area calculations.

Raster structures may lead to increased storage in certain situations, since they store each cell in the matrix regardless of whether it is a feature or simply 'empty' space.

6. a. Explain about GIS data manipulation and analysis.
b. Write short notes on GIS data input and editing. (10+6)

Answer:
(a):



GIS architecture/GIS subsystems

Data Storage and Retrieval: The data storage and retrieval subsystem organizes the data, special and attribute in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the data base. This component usually involves the use of a database management system (DBMS) for maintaining attribute data. Spatial data is usually encoded and maintained in a proprietary file format.

Data Manipulation and Analysis: The data manipulation and analysis subsystem allows the user to define and execute special and attribute procedures to generate derived information. This subsystem is commonly thought of as the heart of a GIS, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.

(b): Data Input: The data input system allows the user to capture, collect, and transform special and thematic data into digital form. The data inputs are usually derived from a combination of hardcopy maps, aerial photographs, remotely sensed images, reports and survey documents.

Data editing: The storage and editing subsystem of GIS provides a valley of tools for storing and maintaining the digital representation of study area. It also provides tools for examining each coverage for mistakes that may have crept into our preparations. The input data that is encoding may consist of a number of errors derived from the original data source as well as errors that have been introduced during the encoding process. There may be errors in co-ordinate data as well as in accuracies and uncertainty in attribute data. Before successfully using the methods of data analysis for any specific application, it is better to intercept errors before they contaminate the GIS database. The process of detecting and removing errors through editing is called **cleaning**.

7. a. Explain the integrated analysis of spatial and attribute data.
b. Describe the basic elements of image interpretation. (10+6)

Answer:
(a): The spatial analysis process
The spatial analysis is composed by a set of chained procedures whose aim is to choose an inferential model that explicitly considers the spatial relationship present in the phenomenon. The initial procedures of analysis include the set of generic methods of exploratory analysis and the visualization of data, in general through maps. These techniques permit the description of the distribution of the variables of study, the identification of observations that are outliers not only in relation to the type of distribution but also in relation to its neighbors, and to look for the existence of patterns in the spatial distribution. Through these procedures it is possible to propose hypothesis about the observations, in a way of selecting the best inferential model supported by the data.

The spatial inferential models are usually presented in three great groups: continuous variation, discrete variation, and the point processes. The resolution of a spatial problem may involve the utilization of one of them or the interaction of some or even all of them. The example below illustrates the differences among these models, how they can be used and how they interact inside the same process where questions, based on real facts, must be responded.
Visceral Leishmaniasis is basically an animal disease but that also affects humans. The dogs are the main domestic reservoirs of the urban disease and there is no treatment for them. The disease is spread by mosquitoes, which reproduce in the soil and in decomposing organic matter, like banana trees and fallen leaves. In the last years there were some epidemic outbreaks in Brazilian cities like Belo Horizontal, Aracatuba, Cuiaba, Teresina, and Natal. The control of the disease is based on the combat against the insect and on the elimination of affected dogs inside the disease focus, an area of 200 meters around the human or canine case.

(b) Basic elements of image interpretation:
(i) Tone: Ground objects of different color reflect the incident radiation differently depending upon the incident wave length, physical and chemical constituents of the objects. The imagery as recorded in remote sensing is in different shades or tones.

(ii) Texture: Texture is an expression of roughness or smoothness as exhibited by the imagery. It is the rate of change of tonal values. Texture can qualitatively be expressed as course, medium and fine. The texture is a combination of several image characteristics such as tone, shadow size, shape and pattern etc., and is produced by a mixture of features too small to be seen individually because the texture by definition is the frequency of tonal changes.

(iii) Association: The relation of a particular feature to its surroundings is an important key to interpretation. Some times a single featured by itself may not be distinctive enough to permit its identification.

(iv) Shape: Some ground features have typical shapes due to the structure or topography. For example air fields and football stadium easily can be interpreted because of their finite ground shapes and geometry whereas volcanic covers, sand, river terraces, cliffs, gullies can be identified because of their characteristic shape controlled by geology and topography.

(v) Size: The size of an image also helps for its identification whether it is relative or absolute. Sometimes the measurements of height (as by using parallax bar) also gives clues to the nature of the object.

(vi) Shadows: shadows cast by objects are sometimes important clues to their identification and interpretation. For example, shadow of a suspension bridge can easily be discriminated from that of cantilever bridge. Similarly circular shadows are indicative of coniferous trees.

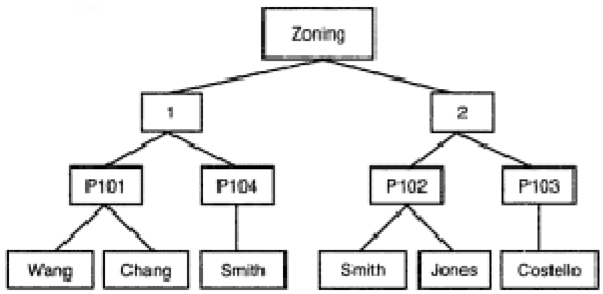
(vii) Size factor or topographic location: Relative elevation or specific location of objects can be helpful to identify certain features. For example, sudden appearance or disappearance of vegetation is a good clue to the underlying soil type or drainage conditions.

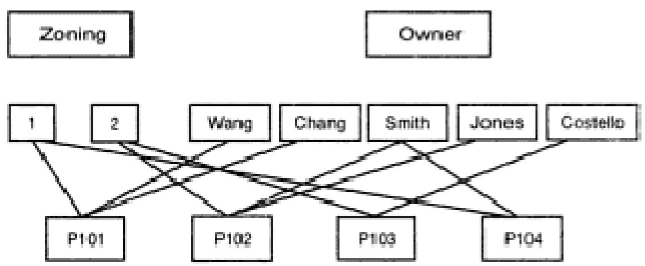
(viii) Pattern: Pattern is the orderly special arrangement of geological topographic or vegetation features. This special arrangement may be two-dimensional or 3-dimensional.

8. a. Explain the types of database design for GIS.
b. Write the quad tree representation. (11+5)

Answer:
(a): Types of database design for GIS: A database is a collection of interrelated tables in digital format. There are at least four types of data base designs that have been proposed in the literature: flat file, hierarchical, network, and relational.

A flat file contains all data in a large table. A feature attribute table is like a flat file. Another example is a spreadsheet with data only. A hierarchical database organizes its data at different levels and uses only the one-to-many association between levels. The simple example is shown in fig.





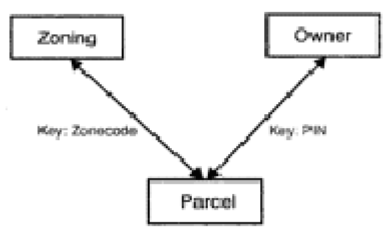


Fig shows us hierarchical levels of zoning, parcel, and owner. Based on the one-to-many association, each level is divided into different branches. A network data builds connections across tables, shown in fig. A common problem with both the hierarchical and network database designs is that the linkages between tables must be known in advance and built into the database at design time. This requirement tends to make a complicated and flexible database and limit the database applications.

A relational database and limit the database applications. GIS vendors typically use the relational model for database, management. A relational database can be connected to each other by keys. A primary key represents one or more attributes whose values can uniquely identify a record in a table. Its counterpart in another table for the purpose of linkage is called foreign key.

(b): Quad tree: The final method of compact storage is a rather difficult approach. Still at least one commercial system called Spatial Analysis System (SPANS), from Tydac, and one experimental system called Quilt are based on this scheme. Like block codes, quadtrees operate on square groups of cells. In this the entire map is successively divided into uniform square groups of cells with the same attribute value. Starting with the entire map as entry points the map is then divided into four quadrants (NW,NE,SW,SE). If any of these quadrants is homogenous containing grid cells with the same value, that quadrant is stored and no further divided into four quadrants, again

NW, NE,SW and SE. Each quadrant is examined for homogeneity. All homogenous quadrants are again stored, and each of the remaining quadrants is further divided and tested in the same way until the entire map is stored, as square groups of cells, each with the same attribute value. In the quadtree structure, the smallest unit of representation is a single grid cell.